Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
ELSEVIER
Journal of Computational and Applied Mathematics 154 (2003) 161–173
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Fourier–Legendre approximation of a probability density from discrete data

Gabriele Inglese

CNR IAC, Via S. Marta 13, Firenze, I-50139, Italy

Received 1 December 2001; received in revised form 16 September 2002

Abstract

We produce a positive approximation of a probability density in $[0,1]$ when only a finite number of values (possibly affected by noise) is available. This approximation is obtained by computing a number of Legendre–Fourier coefficients and applying the Maximum Entropy method. An example of application of this procedure is data-smoothing in the numerical solution of an identification problem for Fokker–Planck equation.

1. Introduction. Position of the problem

We suppose to collect N values of a positive probability density $u:[0,1]\to\mathbb{R}$ in the N distinct points $x_1,\dots,x_N\in(0,1)$. Experimental values can be modelled by the vector $\tilde{u}=(\tilde{u}_k)_{k=1}^N$ defined as

$$\tilde{u}=U(1+\delta u), \tag{1.1}$$

where $U=(u(x_k))_{k=1}^N$ and $\delta u=(\delta u_k)_{k=1}^N$. Assume that $\sum_{k=1}^N(\delta u_k)^2\leqslant\varepsilon^2$. Clearly, $u(x_k)>0$ but we cannot say that $\tilde{u}_k\geqslant 0$.

Remark. It could be more appropriate to regard δu as a vector of N pairwise independent normal random variables with mean equal to zero and assigned variance. Nevertheless, actually this is not relevant at the present stage.

In what follows it is $\|\cdot\|\equiv\|\cdot\|_2$.

E-mail address: gabriele@iaga.fi.cnr.it (G. Inglese).

0377-0427/03/$ - see front matter
PII: S0377-0427(02)00819-1

In order to transform the N-dimensional vector $\tilde{u}$ in a regular function $\tilde{u}^N:[0,1]\to(0,\infty)$ we will carry on the following smoothing procedure:

1. Determine the first N (perturbed) Fourier–Legendre coefficients (FLC) of u from the data vector $\tilde{u}$;
2. Compute the Maximum Entropy solution $\tilde{u}^N$ of the corresponding generalized moment problem;

Then, we estimate stability and accuracy of the approximation produced.

Finally, in Section 3 we verify that $u^N \approx u$ is a suitable input data for numerical solution of an identification problem for Fokker–Planck equation.

2. The smoothing procedure

2.1. Step 1: Fourier–Legendre coefficients

The kth FLC of u is defined as $\lambda_k=\int_0^1 u(x)L_k(x)\,\mathrm{d}x$ where L_k is the Legendre polynomial of degree k shifted in $[0,1]$ with $L_k(1)=1$. We want to compute $\lambda_0,\ldots,\lambda_{N-1}$ from the knowledge of the N values $U_k=u(x_k^{(N)})$. Here, the numbers $x_k^{(N)}$ are the zeros of the $(N+1)$th shifted Legendre polynomial L_N. We have $\sum_{i=1}^N L_{i-1}(x_j^{(N)})\lambda_{i-1}=U_j$ for $j=1,\ldots,N$ i.e.,

$$(V^L)^{\mathrm{T}}\lambda=U.$$

Here, V^L is a Vandermonde-like matrix whose element of position (i,j) is $v_{ij}=L_{i-1}(x_j^{(N)})$ (we recall that A^{T} means the matrix A transposed). It is known that this linear system is quite well conditioned. More precisely there is numerical evidence that the condition number (relative to the Frobenius norm) of V^L is approximately N [7]. In [7] is also proved the following rigorous result: the (Frobenius) condition number of V^L is $\kappa_N^L=\sqrt{\sum_{k=1}^N w_k\sum_{k=1}^N 1/w_k}$ where the w_k's are the Christoffel weights corresponding to the abscissas $x_k^{(N)}$. Straightforward calculations based on [18, formula (15.3.10)] lead to the not optimal (but significant) estimate $\kappa_N^L\leqslant N^{3/2}$.

Suppose that U is affected by noise like in (1.1). In this case $\tilde{\lambda}$ is a vector of *perturbed* LFCs. We have

$$V^L\tilde{\lambda}=\tilde{u}$$

and, after the previous discussion about conditioning of Vandermonde-like matrices, we write the classical estimate

$$\frac{\|\delta\lambda\|}{\|\lambda\|}\leqslant N^2\frac{\|\delta U\|}{\|U\|} \tag{2.1}$$

with $\delta\lambda=\tilde{\lambda}-\lambda$ (see, for example [8] Chapter 1). Alternative methods for obtaining λ are described in Section 4.

2.2. Step 2: Maximum Entropy method

Recall that we are carrying on a smoothing procedure involving a finite set of pointwise measurements. Since our goal is to approximate a continuous probability density u, we have chances to apply Jaynes maximum entropy principle [13] in order to select one among the infinite densities having the same first N Fourier–Legendre coefficients $\lambda_0,\dots,\lambda_{N-1}$ ($\lambda_0=1$). The selected density is "... uniquely determined as the one which is maximally noncommittal with regard to missing information ..." [13].

In practice, we have to solve the following optimization problem:

Minimise the functional $\int_0^1 u(x)\ln u(x)\,\mathrm{d}x$ in the set of the probability densities (i.e., $u\geqslant 0$ and $\int_0^1 u(x)\,\mathrm{d}x=1$) such that $\int_0^1 u(x)L_j(x)\,\mathrm{d}x=\lambda_j$ for $j=0,\dots,N-1$.

It is well known (see, for example [4]) that the minimiser takes the form

$$u^N(x)=\exp\left\{\sum_{k=1}^{N}\alpha_k L_{k-1}(x)\right\},$$

where the coefficients α_k must be determined by solving the system of N nonlinear equations

$$\int_0^1 u^N(x)L_j(x)\,\mathrm{d}x=\lambda_j \tag{2.2}$$

for $j=0,\dots,N-1$.

Existence and uniqueness of the solution $\alpha\in\mathbb{R}^N$ of system (2.2), stated by physical intuition in [13], come from the rigorous analysis of [4, Theorem 3.3; 3]. The rigorous numerical approach to the computation of the Maximum Entropy solution of a finite moment problem [11] can be extended to the determination of u^N. Convergence in the information-theoretic distance $\|u-u^N\|_{IT}=\int_0^1 u(x)\ln u(x)/u^N(x)\,\mathrm{d}x$ is proved in [3]. Uniform convergence is proved in [2].

2.3. Accuracy and stability

Here we deal with the conditioning of system (2.2). First of all we approximate the integral at the l.h.s. by means of a Gaussian quadrature formula and obtain

$$\sum_{i=1}^{N} L_{j-1}(x_i^{(N)})\exp\left\{\sum_{k=1}^{N}\alpha_k L_{k-1}(x_i^{(N)})\right\}w_i^{(N)}+\omega_N^j=\lambda_{j-1} \tag{2.3}$$

for $j=1,\dots,N$. The remainder ω_N^j is expected to decay very fast under suitable bounds on the derivatives of u (see, for example [8] Section 3.2.3 for its expression in terms of the Nth derivative of the integrand).

We introduce the transformation $b(\alpha)$ where

$$b_i(\alpha_1,\dots,\alpha_N)=\exp\left\{\sum_{k=1}^{N}\alpha_k L_{k-1}(x_i^{(N)})\right\}$$

for $i=1,\dots,N$.

If we neglect ω_N^j, Eq. (2.3) appears in vector form (from now on V means V^L) as

$$V^{\mathrm{T}}\mathrm{Diag}[w_i]_{i=1}^{N}b(\alpha)=V^{-1}U.$$

It is not difficult to check that $(V^{\mathrm{T}}\,\mathrm{Diag}[w]V)_{kj}=\int_0^1 L_k(x)L_j(x)\,\mathrm{d}x$ so that $V^{-1}=V^{\mathrm{T}}\mathrm{Diag}[w]$. Hence, we have

$$b(\alpha)=U.$$

A reliable definition of the condition number of a nonlinear system of equations can be found in [16]: Given the equation

$$b(\alpha) = c, \tag{2.4}$$

where α and c are real N-dimensional vectors and $b : \mathbb{R}^N \to \mathbb{R}^N$ is C^∞, the *(relative asymptotic) condition number* κ_N for solving (2.4) at c^0 is

$$\kappa_N = \frac{\|c^0\|}{\|\alpha^0\|} \|[Jb(\alpha^0)]^{-1}\|,$$

where α^0 is the solution of $b(\alpha) = c^0$ and Jb denotes the Jacobian matrix of b.

Remark. Reliability of Rice's definition of conditioning for non linear systems depends on the vectors c^0 and α^0. If these vectors have elements of greatly varying orders of magnitude the condition number as defined can be grossly misleading [Gautschi, private communication]. In our case we are dealing with vectors made up by samples of u and FLC of $\log u$. This kind of vectors can be assumed enough gently behaving without great loss of generality.

It is easy to check that the Jacobian matrix of the transformation b at hand is

$$Jb(\alpha) = \mathrm{Diag}\left[\exp\left\{\sum_{k=1}^{N} \alpha_k L_{k-1}(x_i^{(N)})\right\}\right]_{i=1}^{N} V^{\mathrm{T}}$$

so that

$$\|[Jb(\alpha)]^{-1}\| \leqslant N^{3/2} \max_i \exp\left\{-\sum_{k=1}^{N} \alpha_k L_{k-1}(x_i^{(N)})\right\}. \tag{2.5}$$

To estimate the r.h.s. of (2.5) we make use of a known theorem about the convergence of maximum entropy solutions of finite moment problems:

Theorem BL (Borwein and Lewis [2]). *Let $u \in C^3(0,1)$ be a probability density continuous and strictly positive in $[0,1]$. Let $\lambda \in \mathbb{R}^N$ be the vector of its first N moments $\mu_k = \int_0^1 x^k u(x)\,\mathrm{d}x$ ($k = 0,\dots,N-1$). Let u_N be the Maximum Entropy solution of the corresponding finite moment problem. We have*

$$\lim_{N\to\infty} u^N = u$$

uniformly in $[0,1]$. Moreover

$$\|u^N - u\|^2 \leqslant \frac{|(\log u)'''|}{n^2}.$$

It is clear that if FLCs are available instead of moments, the maximum entropy approximant does not change.

Observe that, after Theorem BL, there exists a $\bar{N}$ such that for any $N > \bar{N}$ we have

$$\min_{[0,1]} \exp\left\{\sum_{k=1}^{N} \alpha_k L_{k-1}(x)\right\} \geqslant \frac{1}{2} \min_{[0,1]} u(x)$$

so that

$$\|[Jb(\alpha^0)]^{-1}\| \leqslant \frac{2N^{3/2}}{\min u} \tag{2.6}$$

for $N \geqslant \bar{N}$.

Remark. In [4, Theorem 3.3] it is proved that if λ is the vector of the first FLCs of some probability density function (PDF), then Maximum Entropy approximant u^N exists and it is unique. If we start from perturbed data $\tilde{u}$ and consequently produce perturbed FLCs $\tilde{\lambda}$, we are not sure that the maximum entropy approximant $\tilde{u}^N$ exists. In fact, the set $F_N = \{(\lambda_1, \dots, \lambda_{N-1}) \in \mathbb{R}^{N-1}$ FLCs of a PDF$\}$ is open, bounded and convex [3] so that we have existence of $\tilde{u}^N$ if ε is small enough. In particular, F^N is the convex hull of the curve $(L_1(t), \dots, L_{N-1}(t))_{t\in[0,1]}$. Moreover, observe that the role of F_N is the same of moment space D^N (see [14]) in finite Hausdorff moment problem. The geometry of D^N has been deeply studied, but very little seems to be known about F_N. So, it is proved that a ball included in D^N (convex hull of the curve $(t, \dots, t^{N-1})_{t\in[0,1]}$) must have diameter lower than $2^{-2(N-2)}$ ([14, Theorem 25.5]), but we can only conjecture that F^N is much less thin (probably "polynomially" thin) as a consequence of the orthogonality of Legendre polynomials.

At this point we can summarize our analysis in the following accuracy-stability estimate:

Theorem 2.1. *Assume that u satisfies the hypothesis of Theorem BL and that* $\min_{[0,1]} u(x) = \gamma > 0$. *Recall that* $\tilde{u} = U + \delta u$ *and* $(\sum_{k=1}^{N} (\delta u_k)^2 \leqslant \varepsilon^2$. *Let* u_N $[\tilde{u}_N]$ *be the Maximum Entropy solution of the generalized moment problem given the FLCs* λ $[\tilde{\lambda}]$ *derived from* U $[\tilde{u}$ *if* ε *is small enough (see Remark)]. In our hypothesis, a positive integer* $\bar{N}$ *exists such that* $\min_{[0,1]} u^N \geqslant \frac{\gamma}{2}$, $(\|u^N\|/\|u\|) \leqslant 2$ *and* $\|u^N - u\| \leqslant \frac{1}{2}$ *when* $N \geqslant \bar{N}$. *Hence, the following estimate holds when* $N \geqslant \bar{N}$:

$$\frac{\|\tilde{u}^N - u\|}{\|u\|} \leqslant \frac{1}{\gamma}\left\{\frac{C}{N} + 2N^{3/2}\varepsilon\right\}. \tag{2.7}$$

Proof. We have

$$\|\tilde{u}^N - u\| \leqslant \|\tilde{u}^N - u^N\| + \|u - u^N\|.$$

Hence,

$$\begin{aligned} \frac{\|\tilde{u}^N - u^N\|}{\|u\|} &\leqslant \frac{\|u^N\|}{\|u\|}\|\tilde{u}^N - 1\| \\ &\leqslant \|\tilde{\alpha} - \alpha\|. \end{aligned}$$

At this point we abuse the concept of (relative asymptotic) condition number linearizing in some sense our problem. We have (in the notation of (2.4)):

$$\begin{aligned} \|\alpha - \tilde{\alpha}\| &\leqslant \|\alpha\|\kappa_N \frac{\|\delta U\|}{\|U\|} \\ &\leqslant \|[Jb(\alpha)]^{-1}\|\|\delta U\| \\ &\leqslant \frac{2N^{3/2}}{\gamma}\varepsilon. \end{aligned}$$

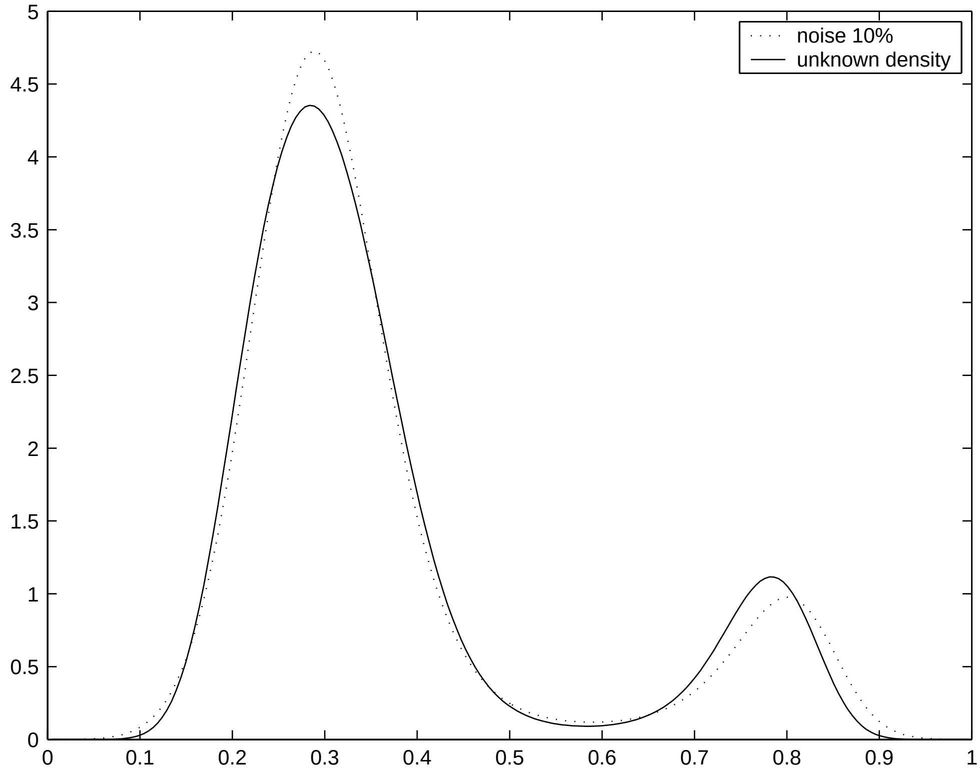

Fig. 1. Reconstruction from nine noisy Fourier–Legendre coefficients.

Finally, from (2.1) we obtain

$$\frac{\|\tilde{u}^{N} - u^N\|}{\|u\|} \leqslant \frac{2N^{3/2}}{\gamma}\varepsilon.$$

As for the discretization error we have

$$\|u - u^N\| \leqslant \frac{\sqrt{|(\log u)'''|}}{2N}$$

so that (2.7) is proved. □

2.4. A numerical example

Results of previous subsection can be visualized by means of a numerical experiment whose output is reported in Fig. 1. Consider the probability density $u(x) = 5.86088\ldots\exp\{-0.027777\ldots(6x-3)^6 - 4.16666\ldots\exp\{-36x^2 + 42x - 12.25\}\}$ in $[0,1]$. Let $u^{(9)}$ be the maximum entropy approximant of Section 2.2 coming from exact values of the FL coefficients $\lambda_0,\ldots,\lambda_8$. The approximation error is $(\|u - u^{(9)}\|/\|u\|) \approx 0.0463$. Let $u_\varepsilon^{(9)}$ come from the perturbed FL coefficients $\tilde{\lambda}_k = \lambda_k(1 + \varepsilon r_k)$ where $k = 0,\ldots,8$, $\varepsilon > 0$ and r_k is a random value in $(-1,1)$. We observed that the error due to magnification of noise (REMN) becomes comparable with the approximation error only for $\varepsilon \approx 0.1$ (see the following table):

$$\varepsilon = 0.01 \rightarrow \text{REMN} = 0.0031$$
$$\varepsilon = 0.05 \rightarrow \text{REMN} = 0.0047$$
$$\varepsilon = 0.10 \rightarrow \text{REMN} = 0.0486$$
$$\varepsilon = 0.50 \rightarrow \text{REMN} = 0.2547$$

Remarks. It is interesting to compare (2.7) with analogous estimates regarding the classical finite moment problem. If we approximate a smooth function u by means of the polynomial $\tilde{u}_N^P$ of best L^2 approximation, calculated from the first N noisy moments $\tilde{\mu}_k = \int_0^1 x^k u(x)\,\mathrm{d}x + \delta u_k$ $(k = 0, \ldots, N-1)$ we get (see, for example [19])

$$\|\tilde{u}_P^N - u\| \leqslant \frac{\|u'\|}{2N} + \exp\{1.72\ldots N\}\varepsilon. \tag{2.8}$$

The same polynomial $\tilde{u}_P^N$ could be obtained from noisy FLCs (see, for example [1]) showing essentially the same stability of (2.7). The maximum entropy solution u^N is anyway preferable in our context because of its positivity. Cancellation of typical Gibbs effect is also observed as an example of regularization after positivity assumption (see for example [6]).

3. An indentification problem for elliptic PDE

Consider the problem of recovering a function V from observations about the evolution in time of the dynamical system

$$\dot{x} = -\frac{\mathrm{d}V}{\mathrm{d}x}. \tag{3.1}$$

Assume that the system is dissipative in the sense that a portion of the intrinsic information of the system itself gets lost along the evolution. It is clear that we have chances to identify attractors, but V (or its derivative f, the so-called drift) is underdetermined.

On the other hand, if we add a Langevin force to (3.1), we obtain the stochastic ODE (see 15,17)

$$\mathrm{d}X_t = -\frac{\mathrm{d}V}{\mathrm{d}x}(X_t) + \sigma \mathrm{d}W_t, \tag{3.2}$$

where $\sigma > 0$ is the *noise strength* of the Langevin perturbation (or diffusion coefficient). A solution of (3.2) (given the initial data $X_0 = x_0$ with probability one) is a Markov stochastic process. We have existence and uniqueness of X_t under suitable hypotheses that make no possible the explosion of the solution:

Theorem H (Has'minskiǐ [10, Theorem 3.4.1]). *Let $K \in \mathbb{R}$ a global attractor for* (3.1). *Assume that there is a bounded, open interval $(a,b) \in \mathbb{R}$ such that $K \subset (a,b)$ and there is a positive real number β such that $(\mathrm{d}V/\mathrm{d}x)(x)x \leqslant -\beta < 0$ for all $x \in \mathbb{R} \backslash (a,b)$. Then there exists a unique Markov process that solves the Cauchy problem* (3.2) *with initial data $X_0 = x_0$ (with probability* 1).

The transition probability densities $p(x,t|x_0,0)$ are well known to solve the Fokker–Planck (FP) diffusion equation

$$\frac{\partial p}{\partial t}(x,t) = -\frac{\mathrm{d}(fp)}{\mathrm{d}x} + \frac{\sigma^2}{2}\frac{\mathrm{d}^2 p}{\mathrm{d}x^2} \tag{3.3}$$

with initial condition $p(x,0) = \delta(x-x_0)$. Morever, [10, Theorem 4.7.1] under suitable additional assumptions we have that $\forall x_0 \in \mathbb{R} \lim_{t\to\infty} p(x,t|x_0,0) = u(x)$ in the sense of probability measures (densities p and u are Radon–Nykodim derivatives) where $u(x) = \mathrm{e}^{-(2/\sigma^2)(V(x)-V(0))}$ is the unique solution of the stationary FP equation

$$-\frac{\mathrm{d}(fp)}{\mathrm{d}x} + \frac{\sigma^2}{2}\frac{\mathrm{d}^2 p}{\mathrm{d}x^2} = 0. \tag{3.4}$$

Now we are ready to state the following problem:

Given σ and u

Determine $V = \int f\,\mathrm{d}x$.

It is easy to check that

$$V(x) = -\frac{\sigma^2}{2}\ln u(x).$$

If one is interested in the drift coefficient, it is $f(x) = -\mathrm{d}V/\mathrm{d}x = -u'/u \approx P_N'(x)$.

Observe that the inverse problem has been formulated in $\mathbb{R}$ while the smoothing procedure in previous section regarded a density in $[0,1]$. Since u is assumed analytical, we could continue it from $[0,1]$ to $\mathbb{R}$. Moreover, we assumed that the system has a bounded global attractor K. Variable x could be rescaled so that $K \subset (0,1)$. In practice, it is preferable that the interesting dynamics is observed mainly in $[0,1]$.

Theorem BL in Section 2.2 suggests to construct a convergent sequence of stable approximations of the potential $V = -\mathrm{d}f/\mathrm{d}x$. Hence, in the light of Section 1 we have

Theorem 3.1. *Let V be analytical in Ω_d (see Theorem BL in Section 2.3) and fulfill hypothesis of theorem H. Then, we have*

$$V(x) = -\lim_{N\to\infty} P_N(x)$$

and

$$f(x) = \lim_{N\to\infty} P_N'(x),$$

where $P_N(x) = \sum_{k=1,N} \alpha_k L_{k-1}(x)$. the numbers $\alpha_1,\dots,\alpha_N$ are obtained from the sample (1.1) *using the method described in Sections* 2.2 *and* 2.3. *Limits are uniform in* $[0,1]$.

4. Variations on Step 1

4.1. Ergodicity

An alternative method for the determination of a finite number of FLCs of u can be formulated when the solution X_t of (3.2) is an *ergodic* Markov process for any initial condition.

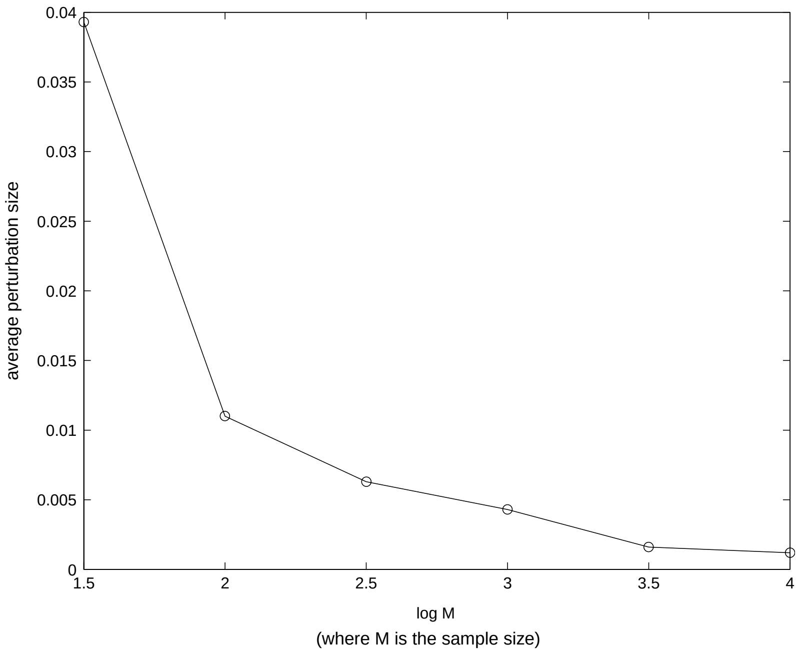

Fig. 2. Montecarlo method for computing FLCs from noisy samples (noise = 10%).

We recall that X_t is said to be ergodic if the following time average limit exists for a function ϕ and equals (w.p. 1) the spacial average with respect to the measure induced by u i.e.,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \phi(X_t)\,\mathrm{d}t = \int_{\mathbb{R}} \phi(x)u(x)\,\mathrm{d}x \tag{4.1}$$

for all measurable functions $\phi : \mathbb{R} \to \mathbb{R}$. Since (4.1) is usually quite difficult to verify directly, we assume that the following ergodicity criterion is verified [10, Section 4.4]:

There exists a bounded interval U such that if $x \in \mathbb{R}\backslash U$ the mean time $E_x\tau$ at which a path issuing from x reaches the set U is finite and $\sup_{x\in K} E_x\tau < \infty$ for all compact set $K \subset \mathbb{R}$.

We introduce the average FLC at time T

$$\lambda_k^T = \frac{1}{T} \int_0^T L_k(X_t)\,\mathrm{d}t$$

so that

$$P\left\{\lim_{T\to\infty} \lambda_k^T = \lambda_k\right\} = 1.$$

It means that w.p. 1 we have that $\forall\varepsilon > 0$ and N there is a T_N such that

$$\sqrt{\sum_{k=0}^{N} (\lambda_k^{T_N} - \lambda_k|^2)} < \varepsilon.$$

Finally, the approximed random FLC $\lambda_k^{T_N}$ $(k = 0,\ldots,N-1)$ can be used as noisy data in step 2 of Section 1.

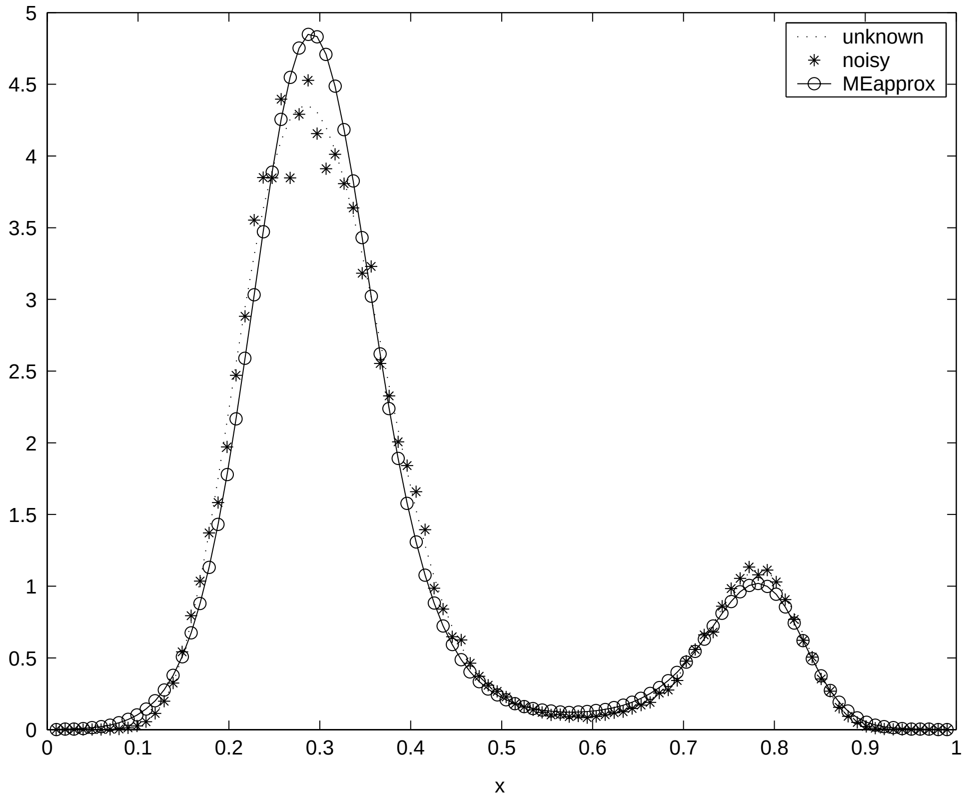

Fig. 3. Maximum entropy approximation of u from a noisy sample (100 points with noise = 10%) via Monte Carlo computation of its first eight FLCs.

4.2. Montecarlo

There is at least another interesting way to manage the vector $\tilde{u}_k = U_k + \delta u_k$ $(k = 1,\dots,M)$ introduced in Section 1 to model the collection of data. We can obtain a number of approximed FLCs of u by way of a Montecarlo-like method without involving interpolation and with chances of reducing arbitrarily the size of the noise affecting the FLCs used for the construction of u^N. The method is based on the following Lemma that generalises a result in [12].

Lemma. *Assume* $\phi \in C^0[0,1]$: $\forall \beta > 0$ *and* $\gamma > 0$ *we have*

$$P\left\{\left|\int_0^1 \phi(x) \sum_{k=1}^{M} \delta u_k \delta(x - x_k^M)\,\mathrm{d}x\right| < \beta\right\} > 1 - \gamma$$

for $M \geqslant \frac{\varepsilon^2}{\gamma\beta^2}\|\phi\|_2$.

Proof. It is sufficient to apply the classical Chebishev inequality (see for example [9, Section 32]). □

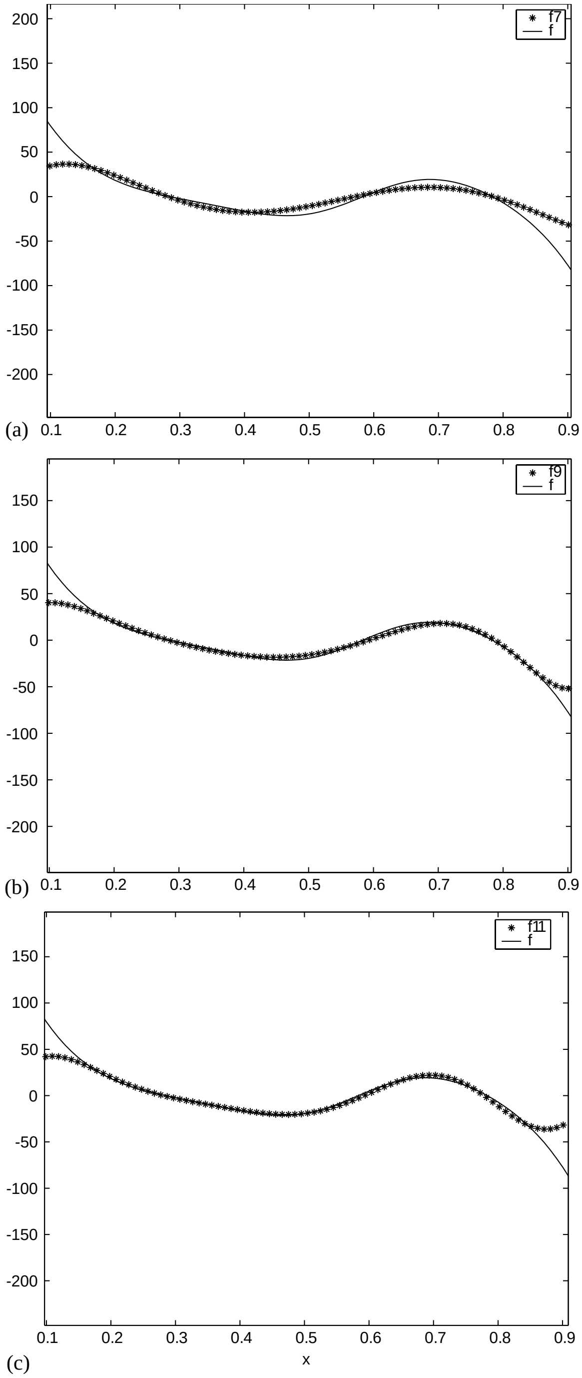

Fig. 4. Reconstruction of a nonlinear drift from data obtained by means of Montecarlo computations.

Since $\forall k = 1, 2, \ldots$ it is $\|L_k\|_2 = 1$, we have

Corollary. *Let u satisfy the hypothesis of Theorem BL and let $\lambda_0, \ldots, \lambda_{N-1}$ be the first FLCs of u. Then, $\forall \beta > 0$ and $\gamma > 0$ we have*

$$P\left\{\left|\sum_{k=1}^{M}(U_k + \delta u_k)L_j(x_k^M) - \lambda_j\right| < \beta\right\} > 1 - \gamma$$

for $M \geqslant \max\{(4\varepsilon^2/\gamma\beta^2), \bar{M}\}$. Here, $\bar{M}$ is such that $|\sum_{k=1}^{\bar{M}}(U_k + \delta u_k)L_j(x_k^{\bar{M}}) - \int_0^1 (U_k + \delta u_k)\delta(x - x_k^{\bar{M}})L_j(x)\,\mathrm{d}x| < \beta/2$.

We can conclude that the effect of a small perturbation of u on the computation of its FLCs goes to zero (with high probability) if $M \to \infty$.

In Fig. 2 we show results obtained in numerical experiments with the density u introduced in Section 2.4.

In Fig. 3 we can observe how much is nonsmooth the perturbed sample $\tilde{u}$. Nevertheless, the ME approximation from data obtained by means of Montecarlo computation actually looks not affected by noise. Finally, in Fig. 4 we show reconstructions of the drift $f = -(6x-3)^5 + 50(6x-3.5)\mathrm{e}^{-(6x-3.5)^2}$ from $\tilde{u}$. It is possible to observe the expected uniform convergence in a subset of $[0,1]$ that do not include zones in which u is very close to zero.

Remark. Since M will be usually very large, we will have in applications $N \ll M$. This fact suggests to consider the extension of properties of rectangular Vandermonde matrices [5] to rectangular Vandermonde-like matrices [Fasino-Inglese, work in progress].

References

[1] U. Amato, M.F. Carfora, The modified moment problem in presence of noise, Approximation Theory Appl. 9 (2) (1993) 50–70.
[2] P. Borwein, A.S. Lewis, Convergence of best entropy estimates, SIAM J. Optim. 1 (2) (1991) 191–205.
[3] B.C. Crain, An information theoretic approach to approximating a probability distribution, SIAM J. Appl. Math. 32 (2) (1977) 339–346.
[4] I. Csiszar, I-divergence geometry of probability distributions and minimization problems, The Annals Probab. 3 (1975) 146–158.
[5] D. Fasino, G. Inglese, On the spectral condition of rectangular Vandermonde matrices, CALCOLO 29 (3–4) (1992) 291–300.
[6] D. Fasino, G. Inglese, Discrete methods in the study of an inverse problem for Laplace's equation, IMA J. Numer. Anal. 19 (1999) 105–118.
[7] W. Gautschi, The condition of Vandermonde-like matrices involving orthogonal polynomials, Linear Algebra Appl. 52–53 (1983) 293–300.
[8] W. Gautschi, Numerical Analysis, An Introduction, Birkäuser, Boston, 1997.
[9] B.V. Gnedenko, The Theory of Probability, MIR, Moscow, 1976.
[10] R.Z. Has'minskiǐ, Stochastic Stability of Differential Equations, Sijthoff and Nordhoff, Alphen, 1980 (Translation of the Russian edition of 1969. Moscow: Nauka).
[11] G. Inglese, A note about the minimum relative entropy solutions of finite moment problems, Numer. Funct. Anal. Optimiz. 16 (9–10) (1995) 1143–1153.

[12] G. Inglese, Identification of the drift coefficient of a Fokker–Planck equation from the moment discretization of its stationary solution, Applicable Anal. 62 (1996) 335–348.
[13] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620–630.
[14] S. Karlin, L.S. Shapley, Geometry of moment spaces, AMS memoirs No. 12, New York, 1953.
[15] B. Oksendal, Stochastic Differential Equations. An Introduction with Applications, Springer, New York, 1991.
[16] J.R. Rice, A theory of condition, SIAM J. Numer. Anal. 3 (1966) 287–310.
[17] H. Risken, The Fokker–Planck Equation, Springer, Berlin, 1989.
[18] G. Szego, Orthogonal Polynomials, AMS Colloq. Publ. Vol. XXIII, New York, 1939.
[19] G. Talenti, Recovering a function from a finite number of moments, Inverse Problems 3 (1987) 501–517.